INFORMATION AND CONTROL 17, 410–415 (1970)

On m-Adic Stochastic Languages

PAAVO TURAKAINEN

Mathematics Department, University of Turku, Finland

The purpose of the paper is to give a necessary and sufficient condition for a certain type of m-adic languages (as defined by Salomaa (1967)) to be non-regular. This solves a problem raised by Salomaa.

I. PRELIMINARIES AND LEMMAS

Let I_m be the alphabet $\{0, 1,..., m-1\}$ $(m \geqslant 2)$ and $I_m{}^*$ the set of all words over I_m including the empty word λ. The length of a word P is denoted by $\lg(P)$. Let φ be a mapping of I_m into $I_m{}^*$. Define

$$\varphi(\lambda) = \lambda,$$

$$\varphi(x_1 \cdots x_k) = \varphi(x_1) \cdots \varphi(x_k),$$

where $x_i \in I_m$. Following Salomaa (1967), we define

$$L(\eta, \varphi) = \{P \in I_m{}^* \mid 0 \,.\, \varphi(P) > \eta\} \qquad (0 \leqslant \eta < 1)$$

where $0 \,.\, \varphi(P)$ is an m-adic expansion, and call $L(\eta, \varphi)$ an *m-adic language.* (It is agreed that $0 \,.\, \varphi(\lambda) = 0$.) The language $L(\eta, \varphi)$ is accepted by the 3-state probabilistic automaton whose initial and final state vectors are $(1, 0, 0)$, $(0, 0, 1)^T$, and transition matrices are

$$M(x) = \begin{bmatrix} m^{\alpha(x)} & 1 - m^{\alpha(x)} - 0 \,.\, \varphi(x) & 0 \,.\, \varphi(x) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where $x = 0, 1,..., m-1$ and $\alpha(x) = -\lg(\varphi(x))$ (cf. Salomaa (1967)).

The alphabet I_m is divided into two separate parts A_m and B_m by defining

$$A_m = \{x \in I_m \mid \varphi(x) \neq \lambda\}, \qquad B_m = \{x \in I_m \mid \varphi(x) = \lambda\}.$$

In what follows, we establish a necessary and sufficient condition for $L(\eta, \varphi)$ to be a nonregular language. To do this, we first denote

$$L_1(\eta, \varphi) = \{P \in A_m^* \mid 0 \,.\, \varphi(P) > \eta\}$$

and begin with three lemmas concerning $L_1(\eta, \varphi)$.

LEMMA 1. *If the cut-point η is a rational number, then $L_1(\eta, \varphi)$ is a regular language.*

Proof. Since η is a rational number, it has an almost periodic m-adic expansion

$$\eta = 0 \,.\, b_1 \cdots b_r a_1 \cdots a_n a_1 \cdots a_n \cdots$$

where $a_i, b_j \in I_m$, $r \geqslant 0$, $n > 0$ and the periodic part does not consist of $(m-1)$'s only.

Let $\{P_1, \ldots, P_s\}$ be the set of all words over A_m such that $\lg(\varphi(P_i)) \leqslant r$ $(i = 1, \ldots, s;\ s \geqslant 0)$. This set is finite, because $\varphi(x) \neq \lambda$ for all $x \in A_m$. Let L_f be the finite language

$$L_f = \{P_i \mid 0 \,.\, \varphi(P_i) > \eta\}.$$

Words $P \in A_m^*$ with $\lg(\varphi(P)) > r$ are divided into classes as follows:

$L_i = \{P \mid \eta$ is of the form $0 \,.\, \varphi(P) \cdots$ and the last letter of $\varphi(P)$ is $a_i\}$ $(i = 1, \ldots, n)$,

$S_1 = \{P \mid \eta$ is not of the form $0 \,.\, \varphi(P) \cdots$ and $0 \,.\, \varphi(P) > \eta\}$,

$S_2 = \{P \mid \eta$ is not of the form $0 \,.\, \varphi(P) \cdots$ and $0 \,.\, \varphi(P) < \eta\}$.

Each word over A_m belongs to exactly one of the classes $\{P_i\}$ $(i = 1, \ldots, s)$, L_j $(j = 1, \ldots, n)$, S_1, S_2. We choose those classes which are not empty. They are the equivalence classes of an equivalence relation E over A_m^*. We show that E is right invariant, i.e., if PEQ then $PR\,EQR$ for every word $R \in A_m^*$. Therefore, assume that PEQ and let $R \in A_m^*$ be arbitrarily fixed. We may assume that $R \neq \lambda$.

If, for some i, $P = P_i$, then $Q = P_i$. Hence, $PR = QR$ and, therefore, $PR\,EQR$.

Assume that, for some i, $P, Q \in L_i$. By the definition of L_i, we have

$$\eta = 0 \,.\, \varphi(P)\, a_{i+1} \cdots a_n a_1 \cdots a_n a_1 \cdots a_n \cdots, \tag{1}$$

$$\eta = 0 \,.\, \varphi(Q)\, a_{i+1} \cdots a_n a_1 \cdots a_n a_1 \cdots a_n \cdots. \tag{2}$$

There are three possibilities, namely, $PR \in L_k$ for some k, $PR \in S_1$, or $PR \in S_2$. If $PR \in L_k$, then, by (1), $\varphi(R) = a_{i+1} \cdots a_k$ or $\varphi(R) = a_{i+1} \cdots a_n (a_1 \cdots a_n)^\nu a_1 \cdots a_k$ where $\nu \geqslant 0$. This, together with (2), implies that η is of the form $0 \,.\, \varphi(QR) \cdots$ where the last letter of $\varphi(QR)$ is a_k. Thus $QR \in L_k$. Since PR and QR are in the same class, $PR\, E\, QR$. If $PR \in S_1$, then η is of the form

$$\eta = 0 \,.\, \varphi(P)\, U \cdots = 0 \,.\, \varphi(Q)\, U \cdots$$

where $\lg(U) = \lg(\varphi(R))$ and $U \neq \varphi(R)$. Moreover, $0 \,.\, \varphi(P)\, \varphi(R) > \eta$. Combining these results, we find that $\varphi(R) > U$. Hence $0 \,.\, \varphi(Q)\, \varphi(R) > 0 \,.\, \varphi(Q) U$. Consequently, $0 \,.\, \varphi(QR) > \eta$. This implies that $QR \in S_1$ and, therefore, $PR\, E\, QR$. In the same way it is verified that $PR\, E\, QR$ whenever $PR \in S_2$.

Next, assume that $P, Q \in S_1$. From the definition of S_1 it follows that $0 \,.\, \varphi(P)\, \varphi(R) > \eta$ and $0 \,.\, \varphi(Q)\, \varphi(R) > \eta$. Thus it is seen that $PR, QR \in S_1$, i.e., $PR\, E\, QR$. In the same way, we find that $PR\, E\, QR$ whenever $P, Q \in S_2$.

By our considerations, the equivalence relation E is right invariant. Moreover, it is of finite index, i.e., the number of equivalence classes is finite. Clearly, $L_1(\eta, \varphi) = L_f + S_1$. In other words, $L_1(\eta, \varphi)$ is the union of some equivalence classes of E. Lemma 1 follows now from Nerode's theorem (cf. Rabin and Scott (1959)).

LEMMA 2. *Assume that η is an irrational number and there exists a number $K > 0$ such that if η is of the form $0 \,.\, \varphi(x_1 \cdots x_i) \cdots$, then $i < K$. Then $L_1(\eta, \varphi)$ is a regular language.*

Proof. From the assumption it follows that there is only a finite number of words $P \in A_m{}^*$ for which η is of the form $0 \,.\, \varphi(P) \cdots$. This implies that the following set of words is finite:

$$X = \{Px \mid P \in A_m{}^*, x \in A_m, \eta \text{ is of the form } 0 \,.\, \varphi(P) \cdots \text{ and } 0 \,.\, \varphi(P)\, \varphi(x) > \eta\}.$$

Lemma 2 follows now from the fact that $L_1(\eta, \varphi) = XA_m{}^*$.

LEMMA 3. *Assume that η is an irrational number and, for each $K > 0$, there exist letters $x_1, \ldots, x_K \in A_m$ for which $\eta = 0 \,.\, \varphi(x_1 \cdots x_K) \cdots$. Then $L_1(\eta, \varphi)$ is a nonregular language.*

Proof. Denote

$$\eta = 0 \,.\, b_1 b_2 b_3 \cdots.$$

Assume, contrary to our assertion, that $L_1(\eta, \varphi)$ is regular. Hence also the language

$$L = \{\varphi(P) \mid P \in L_1(\eta, \varphi)\}$$

is regular, because it is obtained from $L_1(\eta, \varphi)$ by substituting $\varphi(x)$ for each $x \in A_m$. Let DA be a finite (deterministic) automaton accepting the language L. Denote by f, s_1, and n its transition function, initial state, and number of internal states. Let N be the greatest of the numbers $\lg(\varphi(x))$, $x \in A_m$. Let $M = (n+1)N$. For any pair (i, j) such that $1 \leqslant i < j \leqslant M$, let h_{ij} be the smallest positive number satisfying the condition

$$b_{i+h_{ij}} \neq b_{j+h_{ij}}. \tag{3}$$

It exists since, otherwise, the m-adic expansion $0 . b_1 b_2 \cdots$ of η would be almost periodic and, consequently, η is a rational number. Since $1 \leqslant i < j \leqslant M$, it follows that there is only a finite number of numbers h_{ij}. Let H be the greatest of them. By the assumption, there exist letters $x_1, x_2, \ldots, x_{H+M} \in A_m$ such that

$$\eta = 0 . \varphi(x_1 x_2 \cdots x_{H+M}) \cdots.$$

There exist numbers k and l such that $0 < k < l \leqslant n+1$ and

$$f(s_1, \varphi(x_1 \cdots x_k)) = f(s_1, \varphi(x_1 \cdots x_l)). \tag{4}$$

Since $k, l \leqslant n+1$, there exist indices $i, j \leqslant M$ such that b_i and b_j are the last letters of $\varphi(x_1 \cdots x_k)$ and $\varphi(x_1 \cdots x_l)$, respectively. Hence h_{ij} is defined and (3) holds. Since $i + h_{ij} < j + h_{ij} \leqslant H + M$, there exist numbers $\alpha \geqslant 0$, $\beta > 0$ for which $k + \alpha + 1 \leqslant H + M$, $l + \beta \leqslant H + M$, $b_{j+h_{ij}}$ is a letter of $\varphi(x_{l+\beta})$, and

$$\lg(\varphi(x_{l+1} \cdots x_{l+\beta})) = \lg(\varphi(x_{k+1} \cdots x_{k+\alpha})\, U_\alpha), \tag{5}$$

where $U_\alpha \neq \lambda$ is a prefix of $\varphi(x_{k+\alpha+1})$. (If $\alpha = 0$, then the right side is $\lg(U_\alpha)$.) From (5) it follows that $b_{i+h_{ij}}$ is a letter of the word $\varphi(x_{k+1} \cdots x_{k+\alpha})U_\alpha$. Furthermore, its position in this word is the same as that of $b_{j+h_{ij}}$ in the word $\varphi(x_{l+1} \cdots x_{l+\beta})$. This together with (3) implies

$$0 . \varphi(x_{k+1} \cdots x_{k+\alpha})U_\alpha \neq 0 . \varphi(x_{l+1} \cdots x_{l+\beta}). \tag{6}$$

Assume first that, in (6), the number on the right hand side is greater than the number on the left hand side. From this we conclude that

$$0 . \varphi(x_1 \cdots x_k x_{l+1} \cdots x_{l+\beta}) > 0 . \varphi(x_1 \cdots x_k x_{k+1} \cdots x_{k+\alpha})U_\alpha,$$

which gives the result

$$0 \,.\, \varphi(x_1 \cdots x_k x_{l+1} \cdots x_{l+\beta}) > \eta.$$

From the definition of L it now follows that

$$\varphi(x_1 \cdots x_k x_{l+1} \cdots x_{l+\beta}) \in L. \tag{7}$$

It is easy to see that

$$\varphi(x_1 \cdots x_l x_{l+1} \cdots x_{l+\beta}) \notin L. \tag{8}$$

On the other hand, we obtain, by (4),

$$f(s_1 \,,\, \varphi(x_1 \cdots x_k x_{l+1} \cdots x_{l+\beta})) = f(s_1 \,,\, \varphi(x_1 \cdots x_l x_{l+1} \cdots x_{l+\beta})).$$

This contradicts with (7) and (8).

Secondly, assume that the left hand side in (6) is greater than the right hand side. Consequently,

$$0 \,.\, \varphi(x_{k+1} \cdots x_{k+\alpha} x_{k+\alpha+1}) > 0 \,.\, \varphi(x_{l+1} \cdots x_{l+\beta}) U_\beta \,, \tag{9}$$

where U_β is such a prefix of U in the expansion $\eta = 0 \,.\, \varphi(x_1 \cdots x_{l+\beta})U$ (the tail of the expansion is here denoted by U) that the expansions in (9) are of equal length. The following results are now obtained in the same way as (7) and (8):

$$\begin{aligned} &\varphi(x_1 \cdots x_l x_{k+1} \cdots x_{k+\alpha+1}) \in L, \\ &\varphi(x_1 \cdots x_k x_{k+1} \cdots x_{k+\alpha+1}) \notin L. \end{aligned} \tag{10}$$

From (4) we obtain

$$f(s_1 \,,\, \varphi(x_1 \cdots x_k x_{k+1} \cdots x_{k+\alpha+1})) = f(s_1 \,,\, \varphi(x_1 \cdots x_l x_{k+1} \cdots x_{k+\alpha+1})),$$

which contradicts with (10).

The proof of Lemma 3 is now complete.

II. Theorem

Using the above lemmas, we now establish a theorem which generalizes Theorem 4 in (Salomaa, 1967) where it is assumed that the values $\varphi(x)$, $x \in I_m$, are words of equal length.

THEOREM. *The m-adic language $L(\eta, \varphi)$ is nonregular if and only if η is an irrational number and, for every $K > 0$, there exist letters $x_1, \ldots, x_K \in I_m$ such that $\varphi(x_i) \neq \lambda$ $(i = 1, \ldots, K)$ and η is of the form*

$$\eta = 0 \,.\, \varphi(x_1 \cdots x_K) \cdots.$$

Proof. From Lemmas 1, 2, and 3 it follows that the theorem holds for the language $L_1(\eta, \varphi)$. Thus it suffices to prove that $L(\eta, \varphi)$ is regular if and only if $L_1(\eta, \varphi)$ is regular. If $L_1(\eta, \varphi)$ is regular, then also $L(\eta, \varphi)$ is regular, because we obtain it from $L_1(\eta, \varphi)$ by substituting $B_m{}^* X B_m{}^*$ for each letter x. Conversely, if $L(\eta, \varphi)$ is regular, then the same holds for $L_1(\eta, \varphi)$, because we obtain it from $L(\eta, \varphi)$ by substituting the empty word λ for each letter belonging to B_m.

By defining

$$mi(L(\eta, \varphi)) = \{x_1 \cdots x_k \mid 0 \,.\, \varphi(x_k \cdots x_1) > \eta\}$$

we obtain the mirror image of the language $L(\eta, \varphi)$, which is accepted by the 2-state probabilistic automaton whose initial and final state vectors are $(1, 0)$, $(0, 1)^T$ and transition matrices are

$$M(x) = \begin{bmatrix} 1 - 0 \,.\, \varphi(x) & 0 \,.\, \varphi(x) \\ 1 - 0 \,.\, \varphi(x) - m^{\alpha(x)} & 0 \,.\, \varphi(x) + m^{\alpha(x)} \end{bmatrix},$$

where $x = 0, 1, \ldots, m - 1$ and $\alpha(x) = -\lg(\varphi(x))$. It is easy to see that $mi(L(\eta, \varphi))$ is regular if and only if $L(\eta, \varphi)$ is so. Hence we have

COROLLARY. *The above theorem holds for the language $mi(L(\eta, \varphi))$, too.*

This result generalizes Theorem 12 and easily implies Theorem 13 in (Paz, 1966) where, for each $x \in I_m$, the value $\varphi(x)$ equals x.

RECEIVED: January 19, 1970.

REFERENCES

PAZ, A. (1966), Some aspects of probabilistic automata, *Inform. Control* **9**, 26–60.

RABIN, M. O., AND SCOTT, D. (1959), Finite automata and their decision problems, *IBM J. Res. Develop.* **3**, 114–125.

SALOMAA, A. (1967), On m-adic probabilistic automata, *Inform. Control* **10**, 215–219.

Printed in Belgium